

AD-A072 595

NORTH CAROLINA UNIV AT CHAPEL HILL INST OF STATISTICS F/6 12/1
A STUDY OF SEQUENTIAL PROCEDURES FOR ESTIMATING THE LARGER MEAN--ETC(U)
AUG 78 R J CARROLL AFOSR-75-2796

UNCLASSIFIED

MIMEO SER-1189

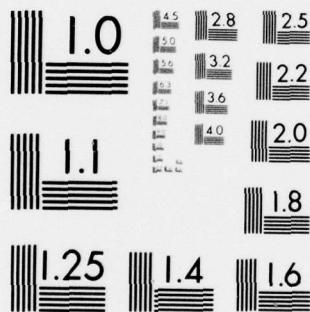
AFOSR-TR-78-1435

NL

| OF |
AD
A072595



END
DATE
FILMED
9-79
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AFCCR-TR- 78-1436

LEVEL ~~II~~

2
b.5.

AD A 072595



A Study of Sequential Procedures for Estimating the Larger Mean

Raymond J. Carroll

Institute of Statistics Mimeo Series #1189

August 1978

DDC FILE COPY

DDC
RECEIVED
AUG 13 1979
D

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

Approved for public release;
distribution unlimited.

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is
approved for public release IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

19 REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 18 AFOSR TR-78-1435	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) 6 A STUDY OF SEQUENTIAL PROCEDURES FOR ESTIMATING THE LARGER MEAN.	9	5. TYPE OF REPORT & PERIOD COVERED Interim rept.
7. AUTHOR(s) 10 Raymond J. Carroll	15	6. PERFORMING ORG. REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of North Carolina Department of Statistics Chapel Hill, North Carolina 27514		8. CONTRACT OR GRANT NUMBER(s) ✓ AFOSR-75-2796 ✓
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332	11	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 16 2304 A5
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 17 A5		12. REPORT DATE August 1978
		13. NUMBER OF PAGES 15
		15. SECURITY CLASS. (of this report) UNCLASSIFIED
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of abstract entered in Block 20, if different from Report) 14 MIMED SER-1189		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Monte-Carlo, Sequential Analysis, Ranking and Selection, Estimating the Larger Mean, Elimination, Mean Square Error		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Blumenthal (1976) considered the problem of sequential estimation of the largest of k normal means when a bound is set on the acceptable mean square error. He showed that his procedure results in only a small savings in sample size when compared to a conservative fixed sample procedure for the case of known variance. Carroll (1978) criticized this procedure because it does not give the user the flexibility of sampling selectively from the k populations. Carroll (1978) defined a procedure which early in the experiment eliminates from further consideration those populations which are obviously not associated with the largest mean and		

DD FORM 1 JAN 73 1473

410064

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract continued.

hence provide little relevant information; his theoretical large-sample calculations indicate possible large savings in sample size with no corresponding increase in mean square error. In this paper we contrast the small sample behavior of the two approaches by means of a Monte-Carlo simulation study; both known and unknown variance are considered.

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DDC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or special
A	

UNCLASSIFIED

LEVEL II

2

A Study of Sequential Procedures for Estimating
the Larger Mean

by

Raymond J. Carroll*
University of North Carolina

Key Words and Phrases: Monte-Carlo, Sequential Analysis, Ranking and
Selection, Estimating the Larger Mean, Elimination, Mean Square Error

* This work was supported by the Air Force Office of Scientific
Research under Contract AFOSR-75-2796.

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

DDC
RECEIVED
AUG 13 1979
RECEIVED
D

1. Introduction

Blumenthal (1976) considered the problem of sequential estimation of the largest of k normal means when a bound is set on the acceptable mean square error. He showed that his procedure results in only a small savings in sample size when compared to a conservative fixed sample procedure for the case of known variance. Carroll (1978) criticized this procedure because it does not give the user the flexibility of sampling selectively from the k populations. Carroll (1978) defined a procedure which early in the experiment eliminates from further consideration those populations which are obviously not associated with the largest mean and hence provide little relevant information; his theoretical large-sample calculations indicate possible large savings in sample size with no corresponding increase in mean square error. In this paper we contrast the small sample behavior of the two approaches by means of a Monte-Carlo simulation study; both known and unknown variance are considered.

2. Known Variance

We are dealing with independent identically distributed observations X_{i1}, X_{i2}, \dots from the i th population, $i = 1, 2$. These are assumed to be normally distributed with means μ_1 and μ_2 and common variance σ^2 . The goal is to estimate the larger mean $\mu_* = \max(\mu_1, \mu_2)$ with a prespecified bound on the mean square error (MSE) r . The asymptotic theorems in Blumenthal (1976) and Carroll (1977) take place as $r \rightarrow 0$. If $\Delta = \max(\mu_1, \mu_2) - \min(\mu_1, \mu_2) = |\mu_1 - \mu_2|$, the mean square error for estimating μ by the larger sample mean based on n observations can be written as

$$\text{MSE} = (\sigma^2/n) H(n^{1/2} \Delta/\sigma) .$$

In order to control the MSE at a prespecified level r when σ is known, Blumenthal (1976) defined the following stopping time.

Definition 2.1. After obtaining m observations from each population, estimate Δ by $|\bar{X}_{1m} - \bar{X}_{2m}| = \hat{\Delta}_m$ and define $\hat{n}(m) = \inf\{n \geq n_0: (\sigma^2/n) H(n^{1/2} \hat{\Delta}_m/\sigma) \leq r\}$ and define

$$N_B = \inf\{m \geq m_0: \hat{n}(m) \leq m\} .$$

Because for $k=2$ populations the risk is a decreasing function of the sample size, one can show that

$$N_B = \inf\{n \geq n_0: \left(\frac{nr}{\sigma^2}\right) \geq H(n^{1/2} \Delta_n/\sigma)\} .$$

Carroll (1978) has shown that Blumenthal's procedure N_B is inefficient in that it does not make use of all the information available in the data. In particular, it does not recognize cases when one population is obviously associated with the smaller mean. Carroll (1978) defined a procedure which attempts to recognize this situation and stop sampling (early in the experiment) for populations which provide information about μ^* . The idea is based on a technique of Swanepoel and Geertsema (1976) and can be described fully as follows. We take $\sigma^2 = 1$ throughout.

Step #1. Choose a small value α , which is the probability of falsely eliminating the population associated with the larger mean. Letting $\phi(\cdot)$ be the standard normal distribution (density) function, define $b = b(\alpha)$ by

$$1 - \phi(b) + b\phi(b) + \phi^2(b)/\phi(b) = \alpha .$$

Step #2. Define a stopping rule

$$N_E = \inf\{n \geq n_0: \hat{\Delta}_n \geq 2^{1/2} ((b^2 + \log n)/n)^{1/2} .$$

Step #3. Define the stopping time $N(\alpha)$ as follows. For a given r , if $[\cdot]$ is the greatest integer function, we will take N_B observations from each population if $N_B \leq N_E$ (no elimination necessary). If $N_E < N_B$ (elimination necessary), we take

N_E observations from the population with smaller mean

$[1/r]+2$ observations from the population with larger mean.

The total sample size is $N(\alpha)$. Note that $N(0.0) = 2N_B$, so Blumenthal's procedure can be read off from the case $\alpha = 0.0$. We chose $n_0 = [\min_x H(x)/r] - 1$.

In order to investigate the small sample performance of $N(\alpha)$, we conducted a Monte-Carlo experiment with 500 iterations and various choices of α, r and Δ .

In Tables 1-4 we record the following information.

- (1) Average value of $N(\alpha)$.
- (2) $N(\alpha)r$
- (3) Bias
- (4) Mean square error divided by r . This should be no more than 1 if we are to meet our goal of controlling MSE by the bound r .

The conclusion one can make from the information in Tables 1-4 is obvious; using elimination results in smaller (sometimes much smaller) sample sizes with no real increase in bias or mean square error.

3. Unknown Variance

For the case that the variance is unknown, the stopping time N_B changes only in that σ^2 is now estimated by

$$s_n^2 = (n-1)^{-1} \sum_{i=1}^n (x_{i1} - x_{i2} - \bar{x}_{1n} + \bar{x}_{2n})^2.$$

The stopping time N_E is again suggested by Swanepoel and Geertsema (1976).

For a given α , we are going to take $n_0 \geq 5$. Define

$$t = .2(1 + a^2/4)^5$$

$$1 - F_4(a) + af_4(a) = \alpha ,$$

where $F_4(f_4)$ is the distribution (density) function of a t distribution with four degrees of freedom. Define

$$h(\alpha, n) = [(tn)^{1/n} - 1]^{\frac{1}{2}} .$$

Then

$$N_E = \inf\{n \geq n_0 : |\bar{X}_{1n} - \bar{X}_{2n}| \geq h(\alpha, n)s_n\} .$$

The results of a Monte-Carlo experiment for this stopping time are given in Tables 5-8.

The conclusion is the same as the case of variance known. Using elimination decreases sample size without materially changing bias or mean square error.

Acknowledgement

The tables were compiled by Mr. Robert Smith whose help is gratefully acknowledged.

References

- Blumenthal, S. (1976). Sequential estimation of the largest normal mean when the variance is known. *Ann. Statist.* 4, 1077-1087.
- Carroll, R.J. (1978). On sequential estimation of the largest normal mean when the variance is known. Institute of Statistics Mimeo Series #1133, University of North Carolina at Chapel Hill. To appear in *Sankhya, Series A*.
- Swanepoel, J.W.H. and Geertsema, J.C. (1976). Sequential procedures with elimination for selecting the best of k normal populations. *S. Afr. Statist. J.* 10, 9-36.

Table 1
Average sample size when the variance is known.

		<u>$\Delta=2.00$</u>	<u>$\Delta=1.00$</u>	<u>$\Delta=.20$</u>
$\alpha = .05$	$r = .10$	18.5	19.5	19.0
	$r = .05$	28.1	35.4	37.2
	$r = .02$	57.7	70.2	91.4
	$r = .01$	107.3	120.4	183.8
$\alpha = .01$	$r = .10$	18.4	19.7	19.0
	$r = .05$	28.9	37.4	37.4
	$r = .02$	58.8	76.6	92.7
	$r = .01$	108.7	127.3	188.4
$\alpha = .00$	$r = .10$	21.1	19.9	19.0
	$r = .05$	41.9	40.2	37.5
	$r = .02$	102.0*	101.4	93.0
	$r = .01$	202.0*	202.0*	189.35

* denotes maximum possible sample size.

Table 2

Average sample size times r when the variance is known.

		<u>$\Delta=2.00$</u>	<u>$\Delta=1.00$</u>	<u>$\Delta=.20$</u>
$\alpha = .05$	$r = .10$	1.85	1.95	1.90
	$r = .05$	1.41	1.77	1.86
	$r = .02$	1.15	1.40	1.83
	$r = .01$	1.07	1.20	1.84
$\alpha = .01$	$r = .10$	1.84	1.97	1.90
	$r = .05$	1.45	1.87	1.87
	$r = .02$	1.18	1.53	1.85
	$r = .01$	1.09	1.27	1.88
$\alpha = .00$	$r = .10$	2.11	1.99	1.90
	$r = .05$	2.10	2.01	1.87
	$r = .02$	2.04	2.03	1.86
	$r = .01$	2.02	2.02	1.89

Table 3
 Bias $\times 10^2$ when the variance is known.

		<u>$\Delta=2.00$</u>	<u>$\Delta=1.00$</u>	<u>$\Delta=.20$</u>
$\alpha = .05$	$r = .10$	1.5	.6	10.6
	$r = .05$.6	.4	5.0
	$r = .02$	-.3	-.3	1.1
	$r = .01$	-.1	-.5	-.3
$\alpha = .01$	$r = .10$.9	1.1	10.7
	$r = .05$.6	.3	5.0
	$r = .02$	-.3	-.3	1.1
	$r = .01$	-.4	-.5	-.2
$\alpha = .00$	$r = .10$.8	1.5	10.7
	$r = .05$.5	.6	5.1
	$r = .02$	-.3	-.3	1.1
	$r = .01$	-.5	-.5	-.2

Table 4

Mean square error divided by r when the variance is known.

		<u>$\Delta=2.00$</u>	<u>$\Delta=1.00$</u>	<u>$\Delta=.20$</u>
$\alpha = .05$	$r = .10$.89	.96	.86
	$r = .05$.88	.92	.76
	$r = .02$.92	.92	.85
	$r = .01$	1.02	1.02	.95
$\alpha = .01$	$r = .10$.91	.99	.87
	$r = .05$.88	.93	.78
	$r = .02$.92	.93	.84
	$r = .01$	1.02	1.02	.94
$\alpha = .00$	$r = .10$.96	1.02	.87
	$r = .05$.93	.97	.78
	$r = .02$.93	.93	.85
	$r = .01$	1.01	1.01	.94

Table 5

Average sample size when the variance is unknown.

		<u>$\Delta=2.00$</u>	<u>$\Delta=1.00$</u>	<u>$\Delta=.20$</u>
$\alpha = .05$	$r = .10$	23.7	25.0	22.8
	$r = .05$	34.5	50.4	47.9
	$r = .02$	64.4	89.5	126.7
	$r = .01$	114.4	138.3	261.8
$\alpha = .01$	$r = .10$	25.4	25.0	22.7
	$r = .05$	41.1	53.4	48.0
	$r = .02$	70.4	109.7	127.3
	$r = .01$	120.4	156.5	263.8
$\alpha = .00$	$r = .10$	25.4	25.0	22.7
	$r = .05$	53.9	53.8	47.9
	$r = .02$	97.8	139.6	127.3
	$r = .01$	146.7	241.5	263.8

* indicates maximum possible sample size obtained.

Table 6

Average sample size times r when the variance is unknown.

		<u>$\Delta=2.00$</u>	<u>$\Delta=1.00$</u>	<u>$\Delta=.20$</u>
$\alpha = .05$	$r = .10$	2.37	2.50	2.28
	$r = .05$	1.72	2.52	2.40
	$r = .02$	1.29	1.79	2.53
	$r = .01$	1.14	1.38	2.62
$\alpha = .01$	$r = .10$	2.54	2.50	2.27
	$r = .05$	2.06	2.67	2.40
	$r = .02$	1.41	2.19	2.55
	$r = .01$	1.20	1.56	2.64
$\alpha = .00$	$r = .10$	2.54	2.50	2.27
	$r = .05$	2.69	2.69	2.40
	$r = .02$	1.96	2.79	2.55
	$r = .01$	1.47	2.42	2.64

Table 7
 Bias $\times 10^2$ when the variance is unknown.

		<u>$\Delta=2.00$</u>	<u>$\Delta=1.00$</u>	<u>$\Delta=.20$</u>
$\alpha = .05$	$r = .10$	-1.9	-2.9	6.4
	$r = .05$.0	.1	2.7
	$r = .02$	-.6	-.1	.7
	$r = .01$	-.2	-.2	-.1
$\alpha = .01$	$r = .10$	-1.9	-2.6	6.7
	$r = .05$.2	-.4	2.7
	$r = .02$	-.6	.5	.7
	$r = .01$	-.2	-.2	-.1
$\alpha = .00$	$r = .10$	-1.7	-2.3	6.9
	$r = .05$	-.3	-.4	2.9
	$r = .02$	-.4	.0	.7
	$r = .01$	-.2	.7	-.1

Table 8

Mean square error divided by r when the variance is unknown.

		<u>$\Delta=2.00$</u>	<u>$\Delta=1.00$</u>	<u>$\Delta=.20$</u>
$\alpha = .05$	$r = .10$.87	.87	.71
	$r = .05$.81	.74	.55
	$r = .02$.96	.88	.62
	$r = .01$.93	.92	.64
$\alpha = .01$	$r = .10$.86	.89	.72
	$r = .05$.78	.70	.55
	$r = .02$.96	.81	.62
	$r = .01$.93	.90	.63
$\alpha = .00$	$r = .10$.86	.89	.73
	$r = .05$.72	.72	.58
	$r = .02$.92	.66	.62
	$r = .01$.93	.77	.63

